Automation using Python

Automation script which accept directory name from user and display all names of duplicate files from that directory.

```
from sys import *
import os
import hashlib

def hashfile(path, blocksize = 1024):
    fd = open(path, 'rb')
    hasher = hashlib.md5()
    buf = fd.read(blocksize)

    while len(buf) > 0:
        hasher.update(buf)
        buf = fd.read(blocksize)

    fd.close()

    return hasher.hexdigest()

def FindDuplicate(path):
    flag = os.path.isabs(path)

    if flag == False:
        path = os.path.abspath(path)

    exists = os.path.isdir(path)

    dups = {}
    if exists:
        for dirName, subdirs, fileList in os.walk(path):
            for filen in fileList:
                path = os.path.join(dirName, filen)
                file_hash = hashfile(path)
                if file_hash in dups:
                    dups[file_hash].append(path)
                else:
                    dups[file_hash] = [path]

        return dups
    else:
        print("Invalid Path")

def PrintDuplicate(dict1):
    results = list(filter(lambda x: len(x) > 1, dict1.values()))

    if len(results) > 0:
        print("Duplicates Found:")

        print("The following files are identical.")

        icnt = 0;
        for result in results:
            for subresult in result:
                icnt+=1
                if icnt >= 2:
                    print('\t\t%s' % subresult)
```

```
57      else:
58          print("No duplicate files found.")
59
60  def main():
61      print("---- Marvellous Infosystems by Piyush Khairnar-----")
62
63      print("Application name : " +argv[0])
64
65      if (len(argv) != 2):
66          print("Error : Invalid number of arguments")
67          exit()
68
69      if (argv[1] == "-h") or (argv[1] == "-H"):
70          print("This Script is used to traverse specific directory and display sizes of files")
71          exit()
72
73      if (argv[1] == "-u") or (argv[1] == "-U"):
74          print("usage : ApplicationName AbsolutePath_of_Directory Extention")
75          exit()
76
77      try:
78          arr = {}
79          arr = FindDuplicate(argv[1])
80          PrintDuplicate(arr)
81
82      except ValueError:
83          print("Error : Invalid datatype of input")
87
88  if __name__ == "__main__":
89      main()
90
```

Output of above script

```
Automation — -bash — 51×20
MacBook-Pro-de-MARVELLOUS:Automation marvellous$ python FileDuplicate.py "ABC"
---- Marvellous Infosystems by Piyush Khairnar-----
Application name : FileDuplicate.py
Duplicates Found:
The following files are identical.
                    /Users/marvellous/Desktop/Automation/ABC/XYZ/DemoABC copy 2.txt
                    /Users/marvellous/Desktop/Automation/ABC/XYZ/DemoABC copy.txt
MacBook-Pro-de-MARVELLOUS:Automation marvellous$
```